



## Information Science and Technology Center Seminar



**Edward R. Dougherty**

Department of Electrical and Computer Engineering, Texas A&M  
University Comp. Bio. Div., Translational Genomics Research Institute  
Depart. Bioinformatics and Comp Bio, U. Texas M.D. Anderson Cancer Center

### "Epistemology of Small-Sample Classification"

**Wednesday, April 14, 2010**

**3:00 - 4:30 PM**

**TA-3, Bldg. 1690, Room 102 (CNLS Conference Room)**

**Abstract:** The accumulation of high-throughput genomic data has spawned a host of proposed gene-expression classifiers to discriminate between phenotypes, in particular, different types, stages, and prognoses for disease. Classical pattern recognition typically involved features possessing contextual meaning, such as geometric features in machine vision and character recognition, and samples that were large in comparison to the number of features. On the other hand, genomic features have generally not depended on biological understanding and the number of features has been extraordinarily large in comparison to sample size. This situation obviously calls for the development of the relevant small-sample theory; however, there has been little effort to understand and address the epistemological issues created by the reversal of the classical paradigm. The consequence is a large number of published papers demonstrably lacking scientific validity and no rigorous scientific road ahead to realize the potential of molecular-based diagnosis and prognosis. This talk discusses the issue of validity in classification, reviews the extensive epistemological failings over the last decade, and proposes an epistemologically sound path ahead based on extending the methods of classical mathematical statistics into the current high-throughput environment.

**Biography:** Edward R. Dougherty is a Professor in the Department of Electrical and Computer Engineering at Texas A&M University in College Station, TX, where he holds the Robert M. Kennedy '26 Chair in Electrical Engineering and is Director of the Genomic Signal Processing Laboratory. He is also co-Director of the Computational Biology Division of the Translational Genomics Research Institute in Phoenix, AZ, and is an Adjunct Professor in the Department of Bioinformatics and Computational Biology at the University of Texas M. D. Anderson Cancer Center in Houston, TX. He holds a Ph.D. in mathematics from Rutgers University and an M. S. in Computer Science from Stevens Institute of Technology, and has been awarded the *Doctor Honoris Causa* by the Tampere University of Technology in Finland. He is a fellow of SPIE, has received the SPIE President's Award, and served as the editor of the SPIE/IS&T Journal of Electronic Imaging. At Texas A&M University he has received the Association of Former Students Distinguished Achievement Award in Research, been named Fellow of the Texas Engineering Experiment Station, and named Halliburton Professor of the Dwight Look College of Engineering. Prof. Dougherty is author of fifteen books, editor of five others, and author of two hundred and fifty journal papers. He has contributed extensively to the statistical design of nonlinear operators for image processing and the consequent application of pattern recognition theory to nonlinear image processing. His current research in genomic signal processing is aimed at diagnosis and prognosis based on genetic signatures and using gene regulatory networks to develop therapies based on the disruption or mitigation of aberrant gene function contributing to the pathology of a disease.